Rationale for nMDS & mMDS

Chapter 5 of the CiMC methods manual describes the operation and rationale of multi-dimensional scaling (MDS) ordination, **Analyse>MDS**. The aim of MDS is to represent the samples as points in low-d space (often 2-d or 3-d, but PRIMER 7 will now compute MDS solutions for any specified range of dimensions), such that the distances apart of all points are as closely matched as possible to the relative dissimilarities (or distances) among the samples, as measured by the resemblance matrix calculated on the (pre-treated) data sheet. The definition of 'closely matched' for the most commonly used form of MDS, **Non-metric MDS (nMDS)**, is that the rank order of dissimilarities among pairs of samples are preserved in ranks of the corresponding pairwise distances in the final ordination plot. The interpretation of a (successful) *n*MDS is therefore straightforward: the closer points are to each other the more similar is their community composition (or suite of environmental data, biomarker responses, particle size distributions, or whatever the variables represent).

PRIMER 7 also provides the more parametric technique of **Metric MDS (mMDS)**, which seeks to interpret the entries in the resemblance matrix as actual distances, so that samples with distance/ dissimilarity *d* are placed at distance *d* in the ordination plot. The key distinction is that the Shepard diagram (a scatter plot of resemblances, *x*, against ordination distances, *y*) is fitted by a straight line in *m*MDS but by a general (non-linear) increasing function in *n*MDS. The much greater flexibility of *n*MDS makes it more suitable for displaying typical community data in low dimensional space, but low-d *m*MDS plots have a useful role to play in ordinations on very few points (as for some means plots) and in region estimates for means (Section 17), or where the resemblance coefficient is, or behaves very like, a genuine distance. Examples might be for normalised Euclidean measures on environmental-type data or community data with low sampling variability from a *short baseline of change* (i.e. relatively little species turnover). Data with more typical sampling variability, but still over a short baseline, can sometimes be well represented in low-d by *threshold metric MDS* (*tm*MDS, Fig. 5.12, CiMC), in which the Shepard plot is fitted by a straight line but not through the origin. Unlike *n*MDS, (*t*)*m*MDS plots thus have a measurement scale interpretable in terms of the resemblances, though all forms of MDS plot are arbitrarily rotatable and reflectable in the axes.

Metric MDS is not Principal Co-ordinate Analysis (PCO), as available in the PERMANOVA+ add-on to PRIMER, though this is a common misconception. PCO is a projection technique from high to low-d, via an eigenvalue decomposition – a generalisation of PCA, see Section 12. The *n*MDS and *m*MDS algorithms are iterative searches, not guaranteed to converge to the optimal solution, hence the need to run them for many random restarts. The default in PRIMER 7 is 50 restarts but if the run time for a single one is not an issue, it is always worth doubling that number, to ensure that a solution is found which is, at least, near-optimal. A working criterion for deciding that enough iterations have been performed is that the same (lowest) stress value is obtained from more than a handful of the restarts. *Stress* measures the scatter in the Shepard plot, i.e. how faithfully the high-d relationships are represented in the low-d ordination – for interpretation of stress values see CiMC.

Revision #1
Created 25 June 2024 00:48:52 by Arden

Updated 25 June 2024 00:58:10 by Arden